[1]

INVERSE PROBLEMS OF SYMBOLIC DYNAMICS

A.YA.BELOV, G.V.KONDAKOV, I.MITROFANOV

Abstract. This paper reviews some results regarding symbolic dynamics, correspondence between languages of dynamical systems and combinatorics. Sturmian sequences provide a pattern for investigation of one-dimensional systems, in particular interval exchange transformation. Rauzy graphs language can express many important combinatorial and some dynamical properties. In this case combinatorial properties are considered as being generated by substitutional system, and dynamical properties are considered as criteria of superword being generated by interval exchange transformation. As a consequence, one can get a morphic word appearing in interval exchange transformation such that frequencies of letters are algebraic numbers of an arbitrary degree.

Concerning multydimensional systems, our main result is the following. Let $P(n)$ be a polynomial, having an irrational coefficient of the highest degree. A word w ($w = (w_n), n \in \mathbb{Z}$) consists of a sequence of first binary numbers of $\{P(n)\}$ i.e. $w_n = [2\{P(n)\}]$. Denote the number of different subwords of w of length k by $T(k)$.

Theorem. *There exists a polynomial $Q(k)$, depending only on the power of the polynomial P, such that $T(k) = Q(k)$ for sufficiently great k.*

1. Introduction

Methods of symbolic dynamics are rather useful in the study of combinatorial properties of words, investigation of problems of number theory and theory of dynamical systems. Let M be a compact metric space, $U \subset M$ be its open subspace, $f : M \to M$ be a homeomorphism of the compact into itself, and $x \in M$ be an initial point. It determines a sequence of points

$$x, f(x), \dots, f^{(n)}(x), \dots$$

With the sequence of iterations, one can associate an infinite binary word

[1]Shanghai University, Moscow Institute of open Education, Moscow Institute of Physicks and Technology, Moscow State University

$$w_n = \begin{cases} a, & f^{(n)}(x_0) \in U \\ b, & f^{(n)}(x_0) \notin U \end{cases}$$

which is called the *evolution* of point x_0. If f is invertible then $n \in \mathbb{Z}$, otherwise $n \in \mathbb{N}$. Symbolic dynamics investigates the interrelation between the properties of the dynamical system (M, f) and the combinatorial properties of the word W_n. For words over alphabets which comprise more symbols, several characteristic sets should be considered: $U_1, \ldots, U_n$. Technical notions regarding combinatorics of words see in section 5.1.

Direct problem of symbolic dynamics consists in the description of properties of the word W, based on the information about dynamic system. *Reverse problem* consists in the description of (M, f, U, x), based on the information about W.

We shall point out some facts, which are known from folklore.

Minimality of the dynamical system corresponds with the uniform recurrence property (see section 5.1).

Uniqueness of invariant measure corresponds with the following property. Let u be a subword of uniformly recurrent (we denote uniformly recurrent as u.r.) word W. For any subword $u \sqsubset W$ lets suppose upper density of occurrence coincides with lower density. The invariant measure is unique, then.

In what cases M is a torus and $f : M \to M$ is its shift? It means that this dynamical system has discrete spectrum. Let W be a superword, obtained by this dynamics. Let T be shift operator. Then mismatch function between W and $T^n(W)$ satisfies following conditions:

(1) There exists sequence $\{n_i\}$ such that $\rho(T^{n_i}(W), W) \to 0$.
(2) There exists co-prime arbitrary large n_i, n_j from this sequence.

We shall analyse these problems. We start from general constructions, regarding torus rotation questions, uniqueness of invariant measure, minimality of dynamical system.

The famous Sturmian sequences and some of their generalizations present situation of "combinatorial paradise". It provides patterns for further investigation. Using language of Rauzy graphs we shall formulate criteria of a superword being generated by interval exchange transformation. Note that any billiard word with rational angles can be obtained via such transformations. (Number of directions of ball is finite and position of ball on the side together with its direction provides phase point, and Phase space is union of some intervals.) On the other hand, using language of Rauzy schemes (obtained from Rauzy graphs by exchanging maximal sequences of vertices of ingoing and

outgoing degree 1 by arches) we get the criteria of the superword to be morphic.

Concerning shifts of multydimensional torus, there is a beautiful theory of Rauzy fractals. For more complicated systems one needs the other patterns for investigation rather then provided by Sturmian sequences.

The rest of this paper is devoted to dynamic systems connected with unipotent transformations of torus. This subject was considered in [19]. Issues related to the study of sequences, obtained by taking fractional part of values of a polynomial at integer points lead to the investigation of such dynamic systems. This problems play an important role in theory of numbers, theory of information transfer and some other branches [1, 2, 3]. Note that unipotent transformation of $\mathbb{T}^2$ has the same relation with circle shift as billiards with arbitrary angles to interval exchange transformations. Sequences appearing in such billiards are analysed in [27].

Inverse problems of symbolic dynamics related to the unipotent transformation of a torus were studied in paper [7] (Unfortunately it was published only in Russian).

We have to point out the paper [19] obtained independently from [7]. Let $Q(X)$ be a real polynomial of degree $d \geq 1$ where the coefficient of X^d is irrational. Define the difference operator $\Delta u_n = u_{n+1} - u_n$ and its iterates $\Delta^2 = \Delta \circ \Delta, \cdots,\ \Delta^d = \Delta \circ \Delta^{d-1}$. The authors establish the theorem according to which the sequence $(\Delta^d \lfloor Q(n) \rfloor)_{n \geq 0}$ takes its value on a 2^d-element alphabet and that

$$p_d(n) = \frac{1}{V(0, 1, \cdots, d-1)} \sum_{0 \leq k_1 < \cdots < k_d \leq n+d-1} V(k_d, \cdots, k_1),$$

where

$$V(k_d, \cdots, k_1) = \prod_{1 \leq i < j \leq d} (k_j - k_i),$$

a Vandermonde determinant. In particular, $p(n)$ depends only on the degree d of the polynomial Q provided the coefficient of X^d is irrational, $p_2(n) = (n+1)(n+2)(n+3)/6$.

In this situation one have to count the number of parts of torus division by images of hyperplane. And proof that points in different regions have different evolutions can be done just by dimension induction because this system provides more information, no theory of quasi-invariant sets and factor dynamics required.

The main theorem of this section is

Theorem 1.1. *There exists a polynomial $Q(k)$, depending only on the degree of the polynomial P, such that $T(k) = Q(k)$ for all sufficiently large k.*

2. Sturmian sequences and their generalizations

The Problems (both direct and inverse) related to the rotation of a circle bring to a class of words which are called *Sturmian words.* Sturmian words are infinite words over a binary alphabet which contain exactly $n+1$ different subwords (factors) of length n for any $n \geq 1$.

Sturmian words provide an example of correspondence between language of dynamical system and combinatorial properties of superwords. We shall formulate classical result:

Theorem 2.1 (Equivalence theorem ([40],[38]).). *Let W be an infinite recurrent word over the binary alphabet $A = \{a, b\}$. The following conditions are equivalent:*

(1) *The word W is a Sturmian word, i.e., for any $n \geq 1$, the number of different subwords of length n that occur in W is equal to $T_n(W) = n + 1$.*
(2) *The word is not periodic and is* balanced*, i.e., any two subwords $u, v \subset W$ of the same length satisfy the inequality $||v|_a - |u|_a| \leq 1$, where $|w|_a$ denotes the number of occurrences of symbol a in the word w.*
(3) *The word $W = (w_n)$ is a mechanical word with irrational α, which means that there exist an irrational α, $x_0 \in [0, 1]$, and interval $U \subset \mathbb{S}^1$, $|U| = \alpha$, such that the following condition holds:*
$$w_n = \begin{cases} a, & {T_\alpha}^n(x_0) \in U \\ b, & {T_\alpha}^n(x_0) \notin U \end{cases}$$
(4) *Word W can be obtained as a limit of the sequence of finite words $\{w_i\}_{i=1}^{\infty}$, such that w_{i+1} can be obtained from w_i via substitution of the following type $a^{k_i}b \to b, a^{k_i+1}b \to a$ or $b^{k_i}a \to a, b^{k_i+1}a \to b$.*

    Iff sequence of these substitutions is periodic, then α is a quadratic irrational.

Sturmian words can be considered as theoretical "paradise" and pattern for further investigations. There are several different ways of generalizing Sturmian words.

First, one can consider *balanced* words over an arbitrary alphabet. Balanced nonperiodic words over an n-letter alphabet were studied in

paper [33] and later in [35]. In the papers [10] and [14], a dynamical system that generates an arbitrary nonperiodic balanced word was constructed.

Secondly, generalization may be formulated in terms of the *complexity function*. Complexity function $T_W(n)$ presents the number of different subwords of length n in the word W. Sturmian words satisfy the relation $T_W(n+1) - T_W(n) = 1$ for any $n \geq 1$. Natural generalizations of Sturmian words are words with minimal growth, i.e., words over a finite alphabet that satisfy the relation $T_W(n+1) - T_W(n) = 1$ for any $n \geq k$, where k is a positive integer. Such words were described in terms of rotation of a circle in paper [15]. Note also that words whose growth function satisfies the relation $\lim_{n\to\infty} T(n)/n = 1$ were studied in paper [16].

Words with complexity function $T_W(n) = 2n + 1$ were studied by P. Arnoux and G. Rauzy ([17, 42, 43]), words with growth function $T_W(n) = 2n + 1$ were analysed by G. Rote [45]. Consideration of the general case of words with linear complexity function involves the study of words generated by interval exchange transformations. The problem of description of such words was posed by Rauzy [43]. Words with linear growth of the number of subwords were studied by V. Berthé, S. Ferenczi, Luca Q. Zamboni ([30], [24]). They investigate combinatorial sequences related with interval exchange transformations. See also works of P.Balázi, Z.Masáková, E.Pelantová ([20],[21], [22]).

3. Interval exchange transformations

Sturmian sequences can be obtained via specific rotation of unit circle. Interval exchange transformation generalizes circle rotation. G.Rauzy posed the question about description of words obtained by interval exchange transformation [43].

S.Ferenci and L.Zamboni [31] obtained the criteria of words generated by interval exchange transformations with following condition: trajectory of every break point don't get on an any break point. In this case obviously a complexity function of the word is equal to $T(n) = (k-1)n + 1$. In fact, this is the answer to the Rauzy question.

In the papers [9, 11] words generated by general piecewise-continuous transformation of the interval were studied. This approach is quite different. The answer to this question is given in terms of the evolution of the *labelled Rauzy graphs* of the word W. The *Rauzy graph* of order k (the *k-graph*) of the word W is the directed graph whose vertices biuniquely correspond to the factors of length k of the word W and the vertex A is connected to vertex B by directed arc iff W has a

factor of length $k+1$ such that its first k letters make the subword that corresponds to A and the last k symbols make the subword that corresponds to B. By the *follower* of the directed k-graph G we call the directed graph $\mathrm{Fol}(G)$ constructed as follows: the vertices of graph $\mathrm{Fol}(G)$ are in one-to-one correspondence with the arcs of graph G and there exists an arc from vertex A to vertex B if and only if the head of the arc A in the graph G is at the notch end of B. The $(k+1)$-graph is a subgraph of the follower of the k-graph; it results from the latter by removing some arcs. Vertices which are tails of (or heads of) at least two arcs correspond to *special factors*; vertices which are heads and tails of more than one arc correspond to *bispecial factors*. The sequence of the Rauzy k-graphs constitutes the *evolution* of the Rauzy graphs of the word W. The Rauzy graph is said to be *labelled* if its arcs are assigned by letters l and r and some of its vertices (perhaps, none of them) are assigned by symbol "-". The *follower* of the labelled Rauzy graph is the directed graph which is the follower of the latter (considered a Rauzy graph with the labelling neglected) and whose arcs are labelled according to the following rule:

(1) Arcs that enter a branching vertex should be labelled by the same symbols as the arcs that enter any left successor of this vertex;
(2) Arcs that go out of a branching vertex should be labelled by the same symbols as the arcs that go out of any right successor of this vertex;
(3) If a vertex is labelled by symbol "–", then all its right successors should also be labelled by symbol "–".

The evolution is said to be *correct* if, for all $k \geq 1$, the following conditions hold when passing from the k-graph G_k to the $(k+1)$-graph G_{k+1} :

(1) Each vertex is an incident to at most two incoming and outgoing arcs;
(2) If the graph contains no vertices corresponding to bispecial factors, then G_{n+1} coincides with the follower $D(G_n)$;
(3) If the vertex that corresponds to a bispecial factor is not labelled by symbol "–", then the arcs that correspond to forbidden words are chosen among the pairs lr and rl;
(4) If the vertex is labelled by symbol "–", then the arcs to be deleted should be chosen among the pairs ll or rr.

The evolution is said to be *asymptotically correct* if this condition is valid for all k beginning with a certain $k = K$. The *oriented* evolution of the graphs means that there are no vertices labelled by symbol "–".

Theorem 3.1 ([11, 9])**.** *A uniformly recurrent word W*

(1) *is generated by an interval exchange transformation if and only if the word is provided with the asymptotically correct evolution of the labelled Rauzy graphs;*
(2) *is generated by an orientation–preserving interval exchange transformation if and only if the word is provided with the asymptotically correct oriented evolution of the labelled Rauzy graphs.*

The proof of this theorem consists of two stages. First proves that these conditions are sufficient for the word to be generated by a piecewise-continuous interval transformation. And the second step is to prove that the sets of uniformly recurrent words generated by piecewise-continuous interval transformations and by the interval exchange transformation are equivalent. In order to do it, there is introduced invariant measure which provides metric: measure of line segment is its length.

4. Substitutional sequences

Important class sequences are so-called *substitutional sequences*. These sequences are invariant under substitution. We refer to the paper [39]. Fibonacci word is such an example. Lets consider substitution $\phi : 0 \to 001, 1 \to 01$. From symbol 0 one can get Fibonacci word as superword $\phi^\infty(0)$. It is Sturmian. Tribonacci word $\tau^\infty(0) = 01020100102010$ can be generated by substitution $\tau(0) = 01, \tau(1) = 02, \tau(2) = 0$.

In fact, in [44] G.Rauzy showed that the Tribonacci minimal subshift (the shift orbit closure of t) is a natural coding of a rotation on the 2-dimensional torus $\mathbb{T}^2$; i.e., is measure-theoretically conjugate to an exchange of three fractal domains on a compact set in $\mathbb{R}^2$. Each domain is translated by the same vector modulo a lattice.

This is one of the most impressive results. It provides description of two dimensional spaces. Symbols correspond with the division of $\mathbb{T}^2$ into fractals, called *Rausy fractals*. Theory of Rauzy fractals was generalized on the so-called *Pisot substitutions*.

This type of dynamical systems provides rich structures with nice picture in multidimensional case. We have correspondence between language of arithmetics, dynamical systems and combinatorics.

In other multidimensional systems one can get some information. Complexity of sequences related with multidimensional systems was studied in [18].

4.1. **language of substitutions and Rauzy graphs.** Consider condition (4) in the theorem 2.1. Sturmian sequence can be obtained as a limit of very specific substitutions. If it is an invariant under some

substitution, then rotation number is quadratic irrationality. Similar fact is known for any rotation of circle. However, there exists substitutions such that eigenvalues of corresponding matrices are algebraic numbers of an arbitrary degree.

The language of Rauzy graphs provides a bridge between combinatorial and topological properties in problems regarding interval exchange transformations. Technique of Rauzy graphs is an important tool in combinatorics of words.

Rauzy graph G_k of Sturmian sequence has one incoming and one outgoing branching vertice. When they coincide, (byspecial word appears) G_{k+1} can be obtained via choosing one of two possibilities, according to the theorem 3.1. This choice corresponds with the decomposition of α in chain fraction. If this choice is made in periodic way α is quadratic irrational.

This fact can be generalized. Let W be a u.r. word. Suppose behavior of Rauzy graphs is periodic in the same sense, then W is equivalent to a superword invariant under some substitution. In order to formulate a theorem one should define what “periodicity of events” in Rauzy graphs means.

Rauzy scheme of W is a sequence of graphs $\{\Gamma_i\}$ such that every vertice of Γ_i is either outgoing or incoming branching vertice of some order, and corresponds to subword of W. Γ_{i+1} can be obtained from Γ_i via exchanging some pathes of length 2 passed through some vertice via arrows and deleting vertices which are not an endpoints of new arrows. *Rauzy scheme is periodic* if there exists $k > 0$ such as for all sufficiently large i there is isomorphism between Γ_i and Γ_{i+k} which fits with transmission from Γ_i to Γ_{i+1}.

Theorem 4.1 ([23])**.** *Uniformly recurrent superword is equivalent to image of some morphism of substitutional invariant sequence iff it has a periodic Rauzy scheme.*

Proof of the implication that if W has a periodic Rauzy scheme it is subtitutional is not so difficult. The main obstacle is the opposite direction. Substitution can be “bad” in the sense that image of letter a_i can have a “parasite” inclusion of image of a_j. So this naive construction fails. We don’t know explicit construction.

Let call *weight of vertice* of Rauzy scheme as a length of corresponding word. In order to prove the theorem, one needs to establish that if W is uniformly recurrent word stable under substitution, then ratios of weights of all Rauzy vertices are bounded. Then one can use J.Cassaigne result [8] saying that if W is uniformly recurrent and $\liminf T_W(n)/n < \infty$, then $\limsup T_W(n+1) - T_W(n) < \infty$. It follows

from the fact that the number of vertices in Rauzy scheme is bounded and gets existence of periodic Rauzy scheme from that. (Condition for morphic uniformly recurrent words $\liminf T_W(n)/n < \infty$ follows from results of Yu.Pritukin [41].

Proof of ratios boundness is based on the following fact. If W is a morphic uniformly recurrent word then there exists a constant $C(W)$ such that for any subword W u occurs in v for any $v \sqsubset W, |v| > C(W) \cdot |u|$. In order to use it, one needs to construct sequence of Rauzy schemes in such a way, that pathes incompatible by inclusion correspond to the words with the same property.

This theorem implies Vershik-Lifshic theorem [46, 47] of periodicity of Bratelli diagrams of Markov compact corresponding to the substitutinal systems.

A *Bratteli diagram* (V, E) is a countable collection V of finite vertex sets, $V = \{V_n\}_{n=1}^{\infty}$ and a countable collection E of finite edge sets $E = \{E_n\}_{n=1}^{\infty}$, along with functions $s : E_n \to V_{n-1}$ and $r : E_n \to V_n$ such that (i) $V_0 = \{\nu_0\}$, (ii) $s : E_n \to V_{n-1}$ and $r : E_n \to V_n$ are onto for all n. We view (V, E) as a directed graph where an edge $e \in E_n$ connects the source vertex $s(e) \in V_n$ to the range vertex $r(e) \in V_{n+l}$. *Periodicity* of Bratelli diagram means that for some k there exists a pair of mappings from V_n to V_{n+k} and from E_n to E_{n+k} preserving functions s and r. With Bratelli diagram one can associate topological dynamics. Details can be founded in [34].

The proof of the Vershik-Lifshic theorem uses straightforward construction. Consider image of $\varphi^{(n)}(a) = (\varphi^{(n-2)})(\varphi^{(2)}(a))$ for some letter a. It consists of blocks corresponding application of $\varphi^{(n-2)}$ to the letters of $\varphi^{(2)}(a)$ and also can be decomposed to the blocks corresponding application of $\varphi^{(n-1)}$ to the letters of $\varphi(a)$. Finite sets forming Bratelli diagrams, consists of sequences of pairs (block, its position in bigger block), they corresponds some subwords of W. On these sets relations of being left and right neighbors inside bigger block can be naturally posed. The first and last occurrence in the bigger block needs special attention, because bigger block itself may have position and can be preceded or followed by another bigger block. Details can be found in [46, 47]. See also [36].

5. Main constructions and definitions

5.1. **Complexity function, special factors, and uniformly recurrent words.** In this section we define the basic notions of combinatorics of words. By L we denote a finite alphabet, i.e., a nonempty

set of elements (symbols). We use the notation A^+ for the set of all finite sequences of symbols or *words*.

A finite word can always be uniquely represented in the form $w = w_1 \cdots w_n$, where $w_i \in A$, $1 \le i \le n$. The number n is called the *length* of word w; it is denoted by $|w|$.

The set A^+ of all finite words over A is a simple semigroup with concatenation as semigroup operation.

If element Λ (the empty word) is included in the set of words, then this is actually the free monoid A^* over A. By definition the length of the empty word is $|\Lambda| = 0$.

A word u is a *subword* (or *factor*) of a word w if there exist words $p, q \in A^+$ such that $w = puq$.

Denote the set of all factors (both finite and infinite) of a word W by $F(W)$. Two infinite words W and V over alphabet A are said to be *equivalent* if $F(W) = F(V)$.

The *beginning* w^b of the word w is a sequence

$$x, f(x), \ldots, f^{(n)}(x), n \ \in \ \mathbb{Z}.$$

We say that symbol $a \in A$ is a *left* (accordingly, *right*) *extension* of factor v if av (accordingly, va) belongs to $F(W)$. A subword v is called a *left* (accordingly, *right*) *special factor* if it possesses at least two left (right) extensions. A subword v is said to be *bispecial* if it is both a left and right special factor at the same time. The number of different left (right) extensions of a subword is called the *left (right) valence* of this subword.

A word W is said to be *recurrent* if each its factor occurs in infinitely many times (in the case of a doubly-infinite word, each factor occurs infinitely many times in both directions). A word W is said to be *uniformly recurrent* or (*u.r word*) if it is recurrent and, for each of its factor v, there exists a positive integer $N(v)$ such that, for any subword u of length at least $N(v)$ of the word W, factor v occurs in u as a subword.

Below we formulate several theorems about u.r words, which will be needed later. The proof of these theorems can be found in monograph [6].

Theorem 5.1. *The following two properties of an infinite word W are equivalent:*

a) For any k there exists $N(k)$ such that any segment of length k of the word W occurs in any segment of length $N(k)$ of the word W;

b) If all finite factors of a word V are at the same time finite factors of a word W, then all finite factors of the word W are also finite factors of the word V.

Theorem 5.2. *Let W be an infinite word. Then there exists a uniformly recurrent word $\widehat{W}$ all of whose factors are factors of W.*

One can consider the action of the shift operator τ on the set of infinite words. *The Hamming distance* between words W_1 and W_2 is the quantity $d(W_1, W_2) = \sum_{n\in\mathbb{Z}} \lambda_n 2^{-|n|}$, where $\lambda_n = 0$ if symbols at the n-th positions of the words are the same and $\lambda_n = 1$, otherwise.

An *invariant subset* is a subset of the set of all infinite words which is invariant under the action of τ. A *minimal closed invariant set*, or briefly, m.c.i.s, is a closed (with respect to the Hamming metric introduced above) invariant subset which is nonempty and contains no closed invariant subsets except for itself and the empty subset.

Theorem 5.3 (Properties of closed invariant sets)**.** *The following properties of a superword W are equivalent:*

(1) *W is a uniformly recurrent word;*
(2) *The closed orbit of W is minimal and is a m.c.i.s.*

Theorem 5.4. *Let W be a uniformly recurrent nonperiodic infinite word. Then*

(1) *All the words that are equivalent to W are u.r. words; the set of such words in uncountable;*
(2) *There exist distinct u.r. words $W_1 \neq W_2$ which are equivalent to the given word and can be written as $W_1 = UV_1$, $W_2 = UV_2$, where U is a left-infinite word and $V_1 \neq V_2$ are right-infinite words.*

6. Unipotent dynamics on a torus

6.1. **Essential evolution of points.** Let $f : M \to M$ be a continuous map on the space M and $U \subset M$ be a subset. The starting point x determines a binary word w describing the evolution as above: $w_n = 1$, if $f^{(n)}(x) \in U$ and $w_n = 0$ if $f^{(n)}(x) \notin U$. We assume that U is an open set, $\mathrm{mes}(\partial U) = 0$ and M is a compact metric space.

Definition 1. *A finite word v^f is said to be an* essential finite evolution of a point x^*, *if every neighborhood of a point x^* contains an open set V, so that for all $x \in V$ holds $v^b_x = v^f$.*

An infinite word w is said to be an essential (infinite) evolution of a point x^*, *if every initial subword is an essential finite evolution of a point x^*.*

The word "evolution" will further denote "essential evolution".

Proposition 6.1. *Let v be a finite word, then the set of points with fixed finite evolution is closed (i.e. consists all its limit points)*

The similar statement holds for an infinite word w. (The intersection of any family of closed sets is closed)

We shall not use the next proposition although it has its own interest:

Proposition 6.2. *Let (M, f, U, x) be a dynamical system without closed invariant subsets, and different points have different evolutions. Let $(\hat{M}, s, U, x)$ be corresponding symbolical dynamical system, i.e. set of all superwords with Tikhonov topology. Then M is naturally isomorphic to factor $\hat{M}$ by spaces consisting of sets of superwords which are essential evolution of one point from M, isomorphism induces natural isomorphisms of dynamical systems.*

6.2. **Morphism of dynamics.**

Definition 2. *A morphism of two dynamics $G : (M_1, f_1) \to (M_2, f_2)$ is a continuous map, such that the diagram*

$$\begin{array}{ccc} M_1 & \xrightarrow{g} & M_2 \\ f_1 \downarrow & & \downarrow f_2 \\ M_1 & \xrightarrow{g} & M_2 \end{array}$$

is commutative.

The notions of *epimorphism, monomorphism* and *isomorphism* are defined in the natural way.

The *factor-dynamics* is naturally defined on the quotient topology iff f permutes the equivalence classes of the map f. Note that the inverse images of points under morphisms are closed.

Definition 3. *A set V is* irreducible, *if its closure is not an inverse image of a closed set under any morphism (except monomorphism) and reducible otherwise.*

Theorem 6.3. *Let a set U be irreducible, then*

1. *Different points have different evolutions.*
2. *For any $\varepsilon > 0$ there exists $N(\varepsilon)$, such that every two words of length $N(\varepsilon)$, corresponding to the initial points on distance greater than ε are different.*

Proof.

1. Classes of points with the same evolution are closed and the map f permutes them.

(2) The proof of this point follows from the proof for the previous one by contradiction and transition to limit.

□

6.3. **Quasi-invariant sets.**

Definition 4. *A dynamics is said to be* minimal *if M does not contain closed invariant sets apart from M and $\emptyset$ and* irreducible*, if it does not contain proper Quasi-invariant sets.*

Definition 5. *A closed set N is* quasi-invariant *if for every two points A and B and a convergent sequence $f^{(n_i)}(A) \to C$ such that $C \in N$ for $n_i \to \infty$ every limit point of the sequence $f^{(n_i)}(B) \in N$.*

Proposition 6.4.

1. *Every closed invariant set is quasi-invariant.*
2. *A partition of quasi-invariant sets corresponds to each factor-dynamics and conversely.*
3. *The image of a quasi-invariant set is quasi-invariant.*
4. *If M is closed invariant set, than one quasi-invariant set uniquely determines the factor–dynamic.*
5. *The set of points with fixed evolution is quasi-invariant.*

Definition 6. $A - B$ cloud (or 0-cloud) with center A, generated by point B *is the set of conditional limit points $f^{(n_i)}(B)$ under the condition $f^{(n_i)}(A) \to A$.*

$A - B$ k-cloud with center A, generated by point B *is the closure of the union of set of conditional limit points $f^{(n_i)}(B)$ under the condition $f^{(n_i)}(A) \to A^*$ where $A^* \in A - B$ $(k-1)$.*

$A - B$ k-cloud with center A, generated by point B *is the closure of the union of the set of $A - B$ k-clouds, $k \in N$.*

Note that $A - B$ k-cloud is closed.

Proposition 6.5. *a) The image of $A - B$ k-cloud under the l-th iteration is $f^{(l)}(A) - f^{(l)}(B)$-clouds.*

b) If $A_n \to A$, $B_n \to B$ then $\rho(A_n - B_n, A - B) \to 0$

Denote $A - B$-cloud by l_0, by L_{i+1} lets denote the closure of the union of all $A_i - B_i$-clouds, for which $A_i,\ B_i \in L_i$. Assume $L_{AB} = \bigcup L_i$. Factorization, generated by L_{AB} is the weakest factorization, that glue points A and B.

6.4. **Unipotent dynamics on a torus.** Problems, connected with the study of the behavior of fractional parts of values of a polynomial at integer points, are in fact the classical problems of symbolic dynamics. Let $P(n)$ be a polynomial of degree $m+1$ with irrational coefficient a_{m+1} of the highest degree. Define a sequence of polynomials $P_k(n)$ $k = 0, \ldots, m$ in the following way:

$$\begin{cases} P_m(n) = P(n), \\ P_{m-1}(n) = P_m(n+1) - P_m(n), \\ \ldots \\ P_{i-1}(n) = P_i(n+1) - P_i(n), \\ \ldots \end{cases}$$

From these formulas it follows, that $P_0(n) = n!a_{m+1}$ is irrational. Put $\varepsilon = P_0(n)$, $x_i(n) = \{P_i(n)\}$ and $x'_i(n) = x_i(n+1)$, then from (1) we obtain the following dynamical system:

$$\begin{cases} x'_m = x_m + x_{m-1} \mod 1 \\ x'_{m-1} = x_{m-1} + x_{m-2} \mod 1 \\ \cdots \\ x'_1 = x_1 + \varepsilon \mod 1, \end{cases} \tag{1}$$

where ε is irrational as $\varepsilon = n!a_{m+1}$. The condition $[2\{P(n)\}] = 0$ turns to the condition $0 \leq x_m(n) < 1/2$.

Consequently the vector $(x'_1, \ldots, x'_m)$ is obtained from $(x_1, \cdots, x_m)$ by unipotent transformation (transformation, corresponding to a linear transformation with unitary eigenvalues).

Images of hyperplanes $x_m = 0$ and $x_m = 1/2$ divide the space on polyhedra. The same word of length k corresponds with the points of the same polyhedron.

6.5. **Mismatch function.**

Definition 7. *A function ρ is said to be a mismatch function of words w and v and is defined in the following way:*

$$\rho(i) = \begin{cases} 0, & if \quad w_i = v_i, \\ 1, & if \quad w_i \neq v_i, \end{cases}$$

A density of mismatch of $\rho(w, v)$ of words w and v is defined by the formula

$$\rho(w, v) = \lim_{i \to \infty} \frac{\sum_{j=1}^{i} \rho(j)}{i}$$

Theorem 6.6. *Let w v be two different evolutions of the point $x_0 \in T$. Then $\rho(w, v) = 0$.*

Proof. From lemma Wail [5] it follows, that the orbit of any point $x_0 \in T$ is everywhere dense and evenly distributed. From the continuity of map f, the condition $\mathrm{mes}(\partial U) = 0$ and the definition of evolution it follows, that $w_n = v_n$ if $f^{(n)}(x_0) \notin \partial U$. The proof of the theorem follows from these statements. □

Theorem 6.7. *Let points x and x^* be different and have different evolutions. Then the density of mismatch $\rho(w_x, w_{x^*})$ is defined and is greater than 0.*

Proof. We can assume that words w_x and w_{x^*} differ in the first position. Consider the direct product $T \times T$. We divide the set of points into two classes: with the same current position and different current position. Let $O \subset T \times T$ be the set of different pairs, then the orbit of pair (x, x^*) lies in O. Closed orbit of any pair is a torus, a minimal closed invariant set, on which the dynamics on torus is realized. Let ρ be the volume of the intersection, then $\rho \neq 0$ and ρ is the density.

□

6.6. **Description of torus factor-dynamics.** We'll consider the dynamics that don't glue the coordinate x_n. Consider the k-th iteration of the transformation of torus:

$$\begin{cases} x_m^{(k)} = x_m + C_k^1 x_{m-1} + \ldots + C_k^m \varepsilon \mod 1 \\ \cdots \\ x_i^{(k)} = x_i + C_k^1 x_{i-1} + \ldots + C_k^i \varepsilon \mod 1 \\ \cdots \\ x_1^{(k)} = x_1 + C_k^1 \varepsilon \mod 1 \end{cases} \tag{2}$$

If the points A $(x_1, \ldots, x_m)$ and B $(x_1 + \Delta x_1, \ldots, x_m + \Delta x_m)$ belong to the set M', then $f^{(k)}(A)$ and $f^{(k)}(B)$ also belong to the same set. From (2) it follows:

$$\begin{cases} \Delta x_m^{(k)} = \Delta x_m + C_k^1 \Delta x_{m-1} + \ldots + C_k^{m-1} \Delta x_1 \mod 1 \\ \cdots \\ \Delta x_i^{(k)} = \Delta x_i + C_k^1 \Delta x_{i-1} + \ldots + C_k^{i-1} \Delta x_1 \mod 1 \\ \cdots \\ \Delta x_1^{(k)} = \Delta x_1 \mod 1 \end{cases} \tag{3}$$

Proposition 6.8. *Closed orbit of any pair of points is a torus T', a minimal closed invariant set, on which the dynamics of torus is realized.*

Remark 1. *For two-dimensional case there exists an area U with analytic boundary and a starting point x^* that have infinitely many different evolutions, that differ in an infinite number of positions.*

Proposition 6.9. *Let 1, ε, Δx_i be linearly independent over $\mathbb{Q}$. Then $A-B$ cloud contains all points which first $i-1$ coincide with coordinates of point B.*

Proof. We may assume that Δx_j is rational when $j<i$ and let $\Delta x_j = p_j/q_j$ be the presentation Δx_j as an irreducible fraction and k is divisible by the product $m!\prod_{j=1}^{i} q_j$, then $x_j^{(kl)} = x_j$ when $j<i$ and systems (2) and (3) can be rewritten as

$$(4)\begin{cases} x_m^{(kl)} = x_m + C_{kl}^1 x_{m-1} + \ldots + C_{kl}^m \varepsilon \mod 1 \\ \cdots \\ x_j^{(kl)} = x_j + C_{kl}^1 x_{j-1} + \ldots + C_{kl}^j \varepsilon \mod 1 \\ \cdots \\ x_1^{(kl)} = x_1 + C_{kl}^1 \varepsilon \mod 1 \\ \Delta x_m^{(kl)} = \Delta x_m + C_{kl}^1 \Delta x_{m-1} + \ldots + C_{kl}^{m-1} \Delta x_1 \mod 1 \\ \cdots \\ \Delta x_j^{(kl)} = \Delta x_j + C_{kl}^1 \Delta x_{j-1} + \ldots + C_{kl}^{j-1} \Delta x_1 \mod 1 (j \geq i) \\ \Delta x_j^{(kl)} = \Delta x_j \mod 1 (j<i) \end{cases}$$

Vector $(x_1^{(kl)}, \ldots, x_m^{(kl)}, \Delta x_i^{(kl)}, \ldots, \Delta x_j^{(kl)})$ by lemma Wail [5] is everywhere dense in the torus of dimension $2m-i+1$, then from definition of $A-B$ cloud it follows that it contains all points, which first $i-1$ coordinates coincide with coordinates of the point B, and other points can be chosen arbitrarily. □

Proposition 6.10. *Let Δx_i be an irrational number. Then there exists a point B_n, the evolution of which coincide with evolution of points A and B, and for which it holds:*

$$\Delta x_i^{B_n} = n \Delta x_i^B.$$

Proof. Lets choose points A and B as B_0 and B_1 respectively. By the method of mathematical induction we'll assume that point B_k is already built, and as for B_{k+1} it suffices to take any conditionally limit point of a sequence $f^{(n_i)}(B_k)$, under the condition $f^{(n_i)}(B_{k-1}) \to B_k$. Note, that the point $B_k \in A-B$ k-cloud.

□

Proposition 6.11. *Let Δx_i be an irrational number. Then there exists a point B_δ, the evolution of which coincide with evolutions of points A*

and B, for which it holds:

$$\Delta x_i^{B_\delta} = \delta, 0 \leq \delta \leq 1.$$

This fact follows directly from the proposition 6.10 and the fact that the set of point with fixed evolution is closed.

Thus the case when Δx_i is irrational reduces to the case when 1, ε, Δx_i are linearly independent over $\mathbb{Q}$.

The case when all Δx_j are rational leads to factor-dynamics where i edge of torus divides in $M_i = \prod_{j=1}^{i} m_j$ parts, where m_j are arbitrary natural numbers and points of the torus $x = x^*$ are identified when for all $1 \leq j \leq m$ holds: $M_j x_j = M_j x_j^*$.

Description of Quasi-invariant sets follows from the proposition 6.11.

Theorem 6.12 (Description of Quasi-invariant sets)**.** *Quasi-invariant set is a shift of abelian group which transforms into themselves under translations by $1/M_i$ along i-coordinate or under all translations along coordinates with the number, greater than some fixed number.*

The theorem provides description of all possible factor-dynamics.

Corollary 6.13. *The class of closed reducible sets consists of sets, which transform into themselves under translations by $1/M_i$ along i-coordinate or under all translations along coordinates with the number, greater than some fixed number. The set $0 \leq x_m < 1/2$ is irreducible.*

6.7. **Proof of the theorem 1.1.** This theorem follows from the next proposition:

Proposition 6.14. *Consider the dynamics of the torus, given by equation (2) and let irreducible set U is given by the condition:*

$$0 \leq x_m \leq 1/2,$$

then there exists such $L(\varepsilon)$ that points with the same finite evolution of length $L(\varepsilon)$ divide torus on closed convex polyhedra, and different polyhedra correspond with the different evolutions.

Proof. The set U is irreducible, so by theorem 6.3 there exists δ' such that $N(\delta')$ evolutions of points at a distance greater than δ' are different. It is obvious, that N-evolutions of all interior points of the polyhedron, obtained after i-th iteration are the same. We call it the evolution of the polyhedron.

Consider parts of the partition, corresponding to words of length $N(\delta')$, then each polyhedron of the partition can't intersect more than

one hyperface from $n+1$ family of planes $f^n[x_1 = q/2,\ q \in N$, otherwise we can choose two points, that belong to this polyhedron and are divided by two planes from $n+1$ family.

Assuming $n+1$ family to be the first and turning the time back we get the situation when $N(\delta')$ evolution of these points doesn't allow to distinguish them, and this is impossible. Polyhedra formed by $n+1$ family of planes with the same evolution can't have common points. Let δ^* be the minimal distance between polyhedra with the same evolutions, then by putting $L(\varepsilon) = \max(N(\delta'),\ N(\delta^*))$ we obtain the number of iterations, from which certainly achieve the equality between number of words of length k and the number of polyhedra in which the torus divides k families of planes $f^i[x_1 = q/2,\ q = 0, \ldots, k-1$. When (ε) is irrational the intersection of more than m hyperplanes is empty and there exists a one-to-one correspondence between points of intersection of m non-parallel planes and polyhedra of partition. By computing the number of points in the intersection of k families of planes $f^i[x_1 = q/2,\ q = 0, \ldots, k-1$ we obtain a polynomial, which defines the number of parts and consequently the number of subwords from some moment. □

The number of points of the intersection of hyperplanes $Q(k)$ and consequently the number of subwords $T(k)$ $(k \geq K)$ of length k can be calculated by the formula:

$$Q(k) = \sum_{0 \leq k_1 < \ldots < k_m \leq k} \begin{vmatrix} 1 & \binom{k_m}{1} & \ldots & \binom{k_m}{m} \\ & & \ldots & \\ 1 & \binom{k_1}{1} & \ldots & \binom{k_1}{m} \end{vmatrix}, \quad \deg Q(k) = \frac{m(m+1)}{2}.$$

For every K there exists P and $k_0 \geqslant K$ such that the equality $T(k) = Q(k)$ holds for $k \geqslant k_0$ and doesn't hold for k_0.

References

[1] Pustylnikov *Sovmestnoe Raspredelenie drobnyh chastei mnogochlena. Uspehi Math Nauk*, 4, Moscow, 1994.

[2] R.N. Izmailov and A.A. Vladimirov. Dimension of aliasing structures. *Journal of Applied Mathematics and stochastic Analysis.*, To appear.

[3] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences.* John Wiley & Sons, N.-Y. etc., 1974.

[4] M. Morse and G.A. Hedlund. Symbolic dynamics II: Sturmian trajectories. *Amer. J. Math.*, 62:1–42, 1940.

[5] H. Weyl. Über der gleichverteilung von zahlen mod. 1. *Math. Ann.*, 77:313–352, 1916.

[6] *A.Ya. Belov, V.V. Borisenko, and V.N. Latyshev*, *Monomial Algebras* // Itogi Nauki i Tekhniki, Sovr. Mat. Pril., Tem. Obzory, vol. 26 (Algebra IV), Moscow, 2002. pp. 35–214 (Russian). NY, Plenum. 1998, pp. 1–173 (English).

[7] *A.Ya. Belov and G.V. Kondakov*, *Inverse Problems of Symbolic Dynamics* // Fundamental and Applied Mathematics, vol. 1, no. 1, 1995, pp. 71–79.

[8] J.Cassaigne. *Special factors with linear subword complexity.* Developments in language theory, II (Magdeburg, 1995), 25-34, World Sci. Publ., River Edge, NJ, 1996.

[9] *A.Ya. Kanel-Belov, A.L. Chernyatiev. Describing the set of words generated by interval exchange transformation.* Comm. in Algebra, Vol. 38, No 7, July 2010, pages 2588–2605.

[10] *A.Ya. Belov and A.L. Chernyatiev*, *Describing Sturmian Words over an n-letter Alphabet* // Math. Met. Appl. IV, MGSU, 1999, pp 122–128.

[11] *Belov A. Ja., Chernyatiev A. L. Words with low complexity and Interval exchange transformations.* Uspehi mat. nauk, Vol. 63, No 1,(379), 2008, pp. 159–160

[12] *Belov A. Ja., Chernyatiev A. L. Words with low complexity and Interval exchange transformations.* Dep. VINITI, No 1048-B2007 from 09.11.07, submitted 09.10.07., 18 pages, ref. 23 items (Russian).

[13] *Ya.G. Sinai*, *Introduction to Ergodic Theory* // Princeton Univ. Press, Princeton, New Jersey, 1977; Moscow, Fazis, 1996.

[14] *A.L. Chernyatiev*, *Balanced Words and Dynamical Systems* // Fundamental and Applied Mathematics, 2007, vol. 13, No 5, pp. 213–224

[15] *A.L. Chernyatiev*, *Words with Minimal Growth Function* // Vestnik Mosk. Gos. Univ., 2008.

[16] *A.Aberkane*, *Words whose complexity satisfies* $\lim p(n)/n = 1$// Theor. Comp. Sci., 307, (2003), 31–46.

[17] *P.Arnoux and G.Rauzy* [1991], *Representation geometrique des suites the complexite* $2n + 1$ // Bull. Soc. Math. France 119, 199–215.

[18] Arnoux, Pierre; Mauduit, Christian; Shiokawa, Iekata; Tamura, Jun-ichi. *Complexity of sequences defined by billiard in the cube.* Bull. Soc. Math. France 122 (1994), no. 1, 1–12.

[19] Arnoux, Pierre; Mauduit, Christian. *Complexite de suites engendrees par des recurrences unipotentes.* (French) [*Complexity of sequences generated by unipotent recurrences.*] Acta Arith. 76 (1996), no. 1, 85–97.

[20] *P.Balázi*, *Infnite Words Coding Three-Interval Exchange* // diploma work CTU (2003).

[21] *P.Balázi*, *Substitution properties of ternary words coding 3- interval exchange*, // ProceedingsWORDS 2003, Eds. T. Harju and J. Karhumaki, (2003), 119–124.

[22] *P. Balázi, Z. Masáková, E. Pelantová*, *Characterization of substitution invariant 3iet words*, submitted to Integers // arXiv:0709.2638, (2007).

[23] *A.Belov, I.Mitrofanov.* Periodicity of Rauzy Schemes for substitutional systems. to appear.

[24] *V.Berthé, S.Ferenczi, and Luca Q.Zamboni* [2007], *Interactions between Dynamics*, Arithmetics and Combinatorics: the Good, the Bad, and the Ugly//

[25] *J.Berstel, P.Séébold*, *Sturmian words*, in: M. Lothaire (Ed.) // Algebraic Combinatorics on Words, Encyclopedia of Mathematics and Its Applications, Vol. 90, Cambridge University Press, Cambridge, 2002 (Chap. 2).

[26] *J.Berstel*, *Resent results on Sturmian words* // Developments in language theory II, 13–24, World Scientific, 1996.

[27] *Cassaigne, J. (F-CNRS-IML); Hubert, P. [Hubert, Pascal] (F-CNRS-IML); Troubetzkoy, S. (F-CNRS-IML) Complexity and growth for polygonal billiards.* Ann. Inst. Fourier (Grenoble) 52 (2002), no. 3, 835–847.

[28] *X.Droubay, J.Justin, G.Pirillo*, *Episturmian words and some construction of de Luca and Rauzy* // Theoret. Comp. Sci.,2000, to appear.

[29] *H.Furstenberg. Poincaré reccurence and number theory* // Bull. Amer. Math. Soc., 5:211–234, 1981.

[30] *S.Ferenczi, Luca Q.Zamboni*, *Combinatorial structure of symmetric k-interval exchange transformations.*

[31] *S.Ferenczi, L.Zamboni*, *Languages of k-interval exchange transformations*, Bull. Lond. Math. Soc. 40 (2008), no. 4, 705–714. http://iml.univ-mrs.fr/∼ferenczi/fz3.pdf

[32] *S.Ferenczi, C.Holton and Luca Q.Zamboni*, *Structure of three-interval exchange transformations II*: a combinatorial description of the trajectories. J. Analyse Math. 89 (2003), p. 239–276.

[33] *R.L.Graham*, *Covering the Positive Integers by disjoints sets of the form* $\{[n\alpha+\beta] : n = 1, 2, \ldots\}$ // J. Combin. Theory Ser A15 (1973) 354–358.

[34] *R.H.Herman, I.F.Putnam and C.F.Skau*, *Ordered Bratteli diagrams, dimension groups and topological dynamics*, Internat. J. Math. 3 (1992), 827–864.

[35] *P.Hubert*, *Well balanced sequences.* // Theoret. Comput. Sci. 242 (2000) 91 pp. 108.

[36] *Application of Adic representations in the investigations of metric, spectral and topological properties of dynamical systems.* Sanct-Petersburg, 1995, 176 pages.

[37] *A.de Luca*, *Sturmian words: structure, combinatorics and their arithmetics* // Theoret. Comp. Sci., 183, (1997), 45–82.

[38] *M.Lothaire*, *Combinatorics on Words*// Encyclopedia of Mathematics and its Applications, Addison-Wesley, Reading, MA, 1983, Vol. 17.

[39] *Mauduit, C. Substitutions, arithmetic and finite automata: an introduction. Substitutions in dynamics, arithmetics and combinatorics,* 35-52, Lecture Notes in Math., 1794, Springer, Berlin, 2002, 37B10 (11B85) PDF Clipboard Series Chapter

[40] *M.Morse and G.A.Hedlund* [1940], *Symbolic dynamics II. Sturmian trajectories*, // Amer. J. Math. 62, 1–42.

[41] *Muchnik, An.A.; Pritykin, Yu.L.; Semenov, A.L. Nearly periodic sequences.* (Russian) Uspekhi Mat. Nauk 64 (2009), no. 5(389), 21–96.

[42] *G.Rauzy*, *Mots infinis en arithmetique*, in: Automata on Infnite Words // Ecole de Printemps d'Informatique Theorique, Le Mont Dore, May 1984, ed. M.Nivat and D.Perrin, Lecture Notes in Computer Science, vol. 192, Springer-Verlag, Berlin etc., pp. 165–171, 1985

[43] *G.Rauzy*, *Exchanges d'intervalles et transformations induites.* (French) Acta Arith. 34, (1979), no. 4, 315–328.

[44] G. Rauzy, *Nombres algebriques et substitutions*, Bull. Soc. Math. France, 110 (1982), p. 147–178.

[45] *G.Rote*, *Sequences with subword complexity* $2n$ J. Number Theory 46 (1994) 196 p. 213.

[46] Vershik, A. M. *The adic realizations of the ergodic actions with the homeomorphisms of the Markov compact and the ordered Bratteli diagrams.* Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 223 (1995), Teor. Predstav. Din. Sistemy, Kombin. i Algoritm. Metody. I, 120–126, 338; translation in J. Math. Sci. (New York) 87 (1997), no. 6, 4054–4058.

[47] Vershik, A. M.; Livshits, A. N. *Adic models of ergodic transformations, spectral theory, substitutions, and related topics. Representation theory and dynamical systems*, 185–204, Adv. Soviet Math., 9, Amer. Math. Soc., Providence, RI, 1992.

[48] *L.Vuillon*, *Balanced words* // Bull. Belg. Math. Soc. Simon Stevin 10 (2003), no. 5, 787–805